Corresponding Author:
Dr. Roberto Fumagalli, Junior Professor, Institute of Philosophy, University of Bayreuth, Germany.
Email: R.Fumagalli@lse.ac.uk

Five Theses on Neuroeconomics

Abstract

Over the last decade, neuroeconomic research has attracted increasing attention by economic modellers and methodologists. In this paper, I examine five issues about neuroeconomic modelling and methodology that have recently been subject to considerable controversy. For each issue, I explicate and appraise prominent neuroeconomists' findings, focusing on those that are claimed to directly inform economic theorizing. Moreover, I assess often-made assertions concerning how neuroeconomic research putatively advances the economic modelling of choice. In doing so, I combine review and critical arguments to provide a methodological evaluation of neuroeconomists' contributions.

Word Count: 9360

Keywords: Neuroeconomics; Choice Models; Economic Theory; Interdisciplinary Integration; Modelling Pluralism.

Introduction

Over the last few years, neuroeconomic research has attracted increasing attention by economic modellers and methodologists. Neuroeconomists' (henceforth, NEs) calls to integrate insights from economics, psychology and neuroscience have sparked intense debates among the practitioners and the philosophers of these disciplines (see e.g. the special issues of *Economics and Philosophy*, 2008, Vol.24, no.3; *Biology and Philosophy*, 2011, Vol.26, no.5; and this Journal, 2010, Vol.17, no.2). At present, neuroeconomics (henceforth, NE) constitutes a highly fragmented discipline, whose relation to economics, psychology and neuroscience is hard to characterize precisely (see e.g. Fumagalli, 2010, and Harrison and Ross, 2010). This, however, implies neither that the hitherto proposed characterizations of NE are equally accurate nor that every contribution at the interface between NE's parent disciplines can be plausibly regarded as progress in NE.

In this paper, I examine five issues about NE modelling and methodology that have recently been subject to considerable controversy, namely: (1) Does NE foster *unification* across its parent disciplines? (2) How do NEs' findings *inform* the economic modelling of choice? (3) Do NEs' findings advance *model selection* in economics? (4) What disciplines provide the *basic constructs* for NEs' accounts of choice? (5) Does NE prompt a revolutionary expansion of the *evidential base* of economic theory? In the five Sections below, I draw on economic methodology, neuro-psychological research and philosophy of science to address these issues in turn. For each issue, I explicate and appraise prominent NEs' findings, focusing on those that are claimed to directly inform economic theorizing. Moreover, I assess often-made assertions concerning how NE research putatively advances the economic modelling of choice. In doing so, I combine review and critical arguments to provide a methodological evaluation of NEs' contributions.

Before proceeding, let me put forward two preliminary caveats. First, different approaches to NE have been distinguished in the literature (see e.g. Craver and Alexandrova, 2008, on 'neuroeconomics proper' and 'economic neural modelling', and Ross, 2008, on 'behavioural economics in the scanner' and 'neurocellular economics'). I shall emphasize the differences between specific approaches whenever these differences are material to the merits of NEs' proposals. For now, this point is worth anticipating. NEs have started to pursue specialized research agendas, and may advance some of these agendas without providing substantive benefits to NE's parent disciplines (see e.g. Kable, 2011, and Vromen, 2010). Still, several leading NEs take their contributions to yield substantive modelling and theoretical benefits to these disciplines. In this paper, I examine whether NEs' contributions actually yield such benefits, devoting particular attention to the economic modelling of choice.[1]

And second, the five issues I address do not exhaust the set of significant questions one may pose about NE (see e.g. Mäki, 2010, on the rhetoric of NE, and Fumagalli, 2013, on NEs' attempts to inform economic welfare analyses). Yet, as I illustrate

[1] NEs frequently use the terms 'models' and 'theory' interchangeably when presenting and discussing their findings. In this paper, I follow this terminological practice unless the cogency of my considerations hinges on specific differences between these two notions (see e.g. Hausman, 1992, ch.5, and Mäki, 1996, sec.5).

below, these five issues target the main respects in which NEs have claimed substantial advances. I take such issues to be especially interesting to economic modellers and methodologists, since they encompass both the major bones of contention between the proponents and the detractors of NE (see e.g. *Sections 2, 3* and *5* on the evidential and explanatory relevance of NE findings for other economists) and the most pressing challenges faced by NEs' attempts to advance economic modelling and theorizing (see e.g. *Section 1* on the intertheoretic relations between NE's parent disciplines, and *Section 4* on the reducibility of economic constructs to neural constructs).

I shall detail in the relevant Sections the interrelations between the five issues I consider and what thesis I advocate regarding each issue. The central theme unifying my five theses can be explicated as follows. The NE literature has grown remarkably over the last decade, and a lot of interesting work is being done by leading NEs. In spite of these *intradisciplinary* advances, NEs have failed to make significant progress with respect to what they formerly singled out as the chief *interdisciplinary* goals of NE research. These failures are by no means unrelated, but originate from the evidential and epistemic challenges inherent in developing a unified modelling and theoretical framework spanning economics, psychology and neuroscience. Some of these challenges may be met thanks to ongoing advances in scanner technology and experimental design. Others, instead, are likely to persist in spite of these advances and hamper progress at the interface between NE's parent disciplines for years to come. In this perspective, the five issues I examine can be seen as interrelated manifestations of a more general impasse, which will continue to hinder NEs' interdisciplinary contributions irrespective of their intradisciplinary achievements.

(1) Does NE foster *unification* across its parent disciplines?

NEs often manifest the ambition to provide a *unified* interdisciplinary framework to model and account for human choice behaviour. For example, Fehr and Rangel allege that NE aims to develop a "detailed computational and neurobiological […] foundation for understanding human behavior across the natural and social sciences" (2011, 4). Similarly, Glimcher maintains that "the goal of [NE] is to produce a single unified model of human decision making that spans the economic, psychological, and neuroscientific levels of analysis" (2011, 4). NEs do not presently share precise views as to what contributions NE's parent disciplines are to provide to their still-to-come 'single unified model' (see *Sections 2-4*). Suppose, for the sake of argument, that they reached consensus on this issue. Even so, two major questions would remain regarding NEs' calls to develop such a single unified model. The first question is whether integrating findings and modelling tools from disciplines as diverse as economics, psychology, and neuroscience into a single unified model is *feasible*. The second question is whether this unification, even if feasible, brings valuable modelling and theoretical *benefits* to NE's parent disciplines. I already commented on the latter issue in other articles (see e.g. Fumagalli, 2011). In this Section, I critically assess the prospects of NEs' unificationist efforts.

Over the last few decades, several attempts have been made to integrate findings and modelling tools across economics and psychology (see e.g. Kahneman and Tversky, 1979, on prospect theory, and Loomes and Sugden, 1982, on regret theory). Fruitful interdisciplinary interactions have taken place in various areas of neuro-psychological research (see e.g. Mundale and Bechtel, 1996, and Piccinini and Craver, 2011, on the integration of functional and mechanistic insights across cognitive psychology and computational neuroscience). According to leading NEs, promising advances are underway also at the interface between economics and neuroscience. These alleged advances are prompted not just by bottom-up neuroscientific findings, but also by top-down behavioural evidence.

To give one example, after observing that neural activations in the lateral intraparietal area correlate with the relative expected desirability of saccadic eye movements in monkeys, Dorris and Glimcher conjecture that activations in this area "may also encode the subjective desirability of actions in humans" (2004, 376). Suppose this was actually the case. As noted by Glimcher (2011, 234-6), it would be mistaken to infer from this result that the human brain encodes *only* relative expected subjective values. For any organism whose valuation system stores only these values would make intransitive choices much more often than humans. In this respect, observed behavioural patterns constrain NE models by suggesting that both relative and absolute value signals must be encoded in the human neural architecture.

This example nicely illustrates how economic and neuro-psychological findings can constrain NE models of choice. However, NEs' integrative achievements are typically confined to specific modelling contexts and do not directly advance the construction of a single unified model spanning NE's parent disciplines. Moreover, there are several reasons to doubt the prospects of NEs' attempts to develop this single unified model. I shall comment on two such reasons in turn.

The first reason relates to the *dissimilar interpretations* that NEs and other economists respectively give to choice models. The following contrast highlights a profound divide between these interpretations (see e.g. Kable and Glimcher, 2009, on 'traditional' economic models and 'because' NE models; see also Fehr and Rangel, 2011, on 'as if' and 'as is' models of choice). On the one hand, economic modellers usually abstract from the mechanistic underpinnings of choice and do not take a position as to what neuro-psychological processes underlie observed decisions. On the other hand, several NE modellers aim to show that agents who behave as predicted by standard economic theory do so because specific neural areas encode the relative expected values of the available options. The fact that economic modellers do not take a position as to what neuro-psychological processes underlie observed decisions does not *per se* vindicate these modellers' interpretation of choice models. For neural findings may advance the economic modelling of choice even though NEs and other economists give dissimilar interpretations to choice models (see e.g. Camerer, 2008, on calls to use neural findings to discriminate between competing economic models, and Glimcher, 2011, ch.4-6, on attempts to build more predictive and explanatory models of choice). Still, the diversity of these interpretations poses severe pragmatic and

methodological challenges to NEs' calls to use neural findings in constructing and evaluating economic models.[2]

My second reason to doubt the prospects of NEs' unificationist efforts builds on NEs' divergences regarding the *ultimate goals* of NE research. At first glance, most NE studies seem primarily concerned with investigating the neural substrates of choice.[3] This, however, falls short of implying that NEs share the same view as to what NE ultimately aims to accomplish. To see this, let us contrast the main research goals that are respectively associated with so-called 'behavioural economics in the scanner' and 'neurocellular economics' (Ross, 2008, 473-4). On the one hand, the proponents of behavioural economics in the scanner (e.g. Camerer et al., 2005) employ the neural evidence obtained in specific experimental tasks to discriminate between competing economic models and shed light on specific violations of standard economic theory. On the other hand, the advocates of neurocellular economics (e.g. Glimcher, 2011, ch.3-5) use economists' constrained optimization techniques to model the workings of the neural substrates of decisions and improve models of reward valuation in cognitive and computational neuroscience. These divergences do not preclude distinct groups of NEs from pursuing some common research goals (see e.g. Vromen, 2011, on the development of more predictive models of choice). Still, they point to a profound methodological cleavage, which has pervasively shaped the research agendas of prominent NEs (see e.g. Glimcher et al., 2005).

A proponent of NE may concede that NEs' divergences concerning the ultimate goals of NE research hamper unification across NE's parent disciplines. At the same time, she might object that the practitioners of these disciplines could build integrated models of decision-making, and yet continue to investigate their phenomena of interest by means of traditional methodologies and modelling approaches. The idea is that economists, psychologists and neuroscientists develop a common NE framework for modelling choices that transcends the methodologies and the modelling approaches entrenched in each discipline (Glimcher and Rustichini, 2004, 452). In this way, NE could allegedly improve its parent disciplines on their own terms without having to literally unify these disciplines' methodologies and modelling approaches (Camerer, 2008, 59).

NEs' calls for interdisciplinary unification seem less implausible if one charitably takes NEs to be developing - not so much a single unified model of decision-

[2] I expand in *Section 3* on the challenges faced by NEs' calls to employ neural findings for model selection purposes in economics. As to the predictive and explanatory benefits yielded by NEs, I articulated in previous works two challenges to NEs' calls to use neural findings in constructing and evaluating economic models. The first challenge (Fumagalli, 2011) is that due to the trade-offs between the desiderata respectively valued by NEs and other economists, showing that neural findings help economists to satisfy specific desiderata falls short of implying that NEs enable economists to build better models of choice. The second challenge (Fumagalli, 2014) is that NEs have hitherto failed to significantly improve economic models even with regard to individual desiderata.

[3] I speak of 'neural substrates of choice' broadly to indicate both algorithmic and neuro-anatomical findings. The proponents of NE frequently emphasize that NE studies target both sets of findings. For instance, Montague argues that there are "two natural [NE]", one which investigates "the behavioural algorithms running on [the] neural tissue", the other which examines "the way that neural tissue is built, sustains itself through time, and processes information efficiently" (2007, 219).

making, but rather - a common modelling framework, i.e. a collection of models which share a relatively precise characterization of NE, pursue similar explanatory aims, and reflect consistent views of the relationship between NE's parent disciplines. However, the differences between distinct approaches to NE (see e.g. *Sections 2* and *4*) hamper NEs' attempts to provide a common modelling framework spanning economics, psychology and neuroscience. Moreover, NEs presently lack the shared basic constructs to develop such framework. To be sure, some constructs (e.g. utility) do figure in various models of choice across NE's parent disciplines. Still, these constructs are given rather different interpretations by the practitioners of those disciplines, and NEs often fail to acknowledge how heavily such differences hinder their unificationist ambitions (see e.g. Fumagalli, 2013, on distinct notions of utility). Paraphrasing what Kahneman observed concerning economics and psychology, "there are no immediate prospects of economics [psychology and neuroscience] sharing a common theory of human behavior" (2003, 165-6).

More generally, the point remains that economists, psychologists and neuroscientists have made considerable advances by relying on highly specialized methodologies and modelling approaches. This historical record does not license the isolationist attitude exhibited by some detractors of NE (see e.g. Gul and Pesendorfer, 2008). Yet, coupled with the challenges faced by NEs' attempts to build a unified interdisciplinary model of choice, it counsels choice modellers to reflect before embarking on ambitious transdisciplinary Russian campaigns. That is to say, one welcomes NEs' attempts to integrate findings and modelling tools across different decision sciences. Even so, severe concerns remain regarding the project to provide a single unified model of choice spanning economics, psychology and neuroscience. In this perspective, NEs' calls for interdisciplinary unification seem inspired more by unreflective enthusiasm for recent advances in neuro-psychological research than by principled observations concerning how NE informs its parent disciplines.

(2) How do NEs' findings *inform* the economic modelling of choice?

NEs advocate heterogeneous positions concerning the ways in which their findings are supposed to inform NE's parent disciplines. Two positions are frequently contrasted with regard to how such findings putatively inform the economic modelling of choice.[4] On the one hand, *incremental* NE "adds variables to conventional accounts of decision making or suggests specific functional forms to replace 'as if' assumptions that have never been well supported empirically" (Camerer et al., 2005, 10). On the other hand, *radical* NE asks "how economics might have evolved differently if it had been informed from the start by insights and findings now available from neuroscience" (ibid., 10). These claims, which echo former categorizations of behavioural economists' contributions (see e.g. Rabin, 1998), highlight what is commonly regarded as a fundamental divide between distinct approaches to NE research (see e.g. Gul and Pesendorfer, 2008).

[4] Expressions such as 'informing economic models' and 'informing the economic modelling of choice' are used in different senses by distinct authors. In this paper, I employ such expressions to refer to the thesis that NE findings have direct evidential and explanatory relevance for the economic modelling of choice.

In this Section, I assess various interpretations of those claims and expand on an alternative conceptualization of the incremental/radical divide.

Taken literally, Camerer et al.'s characterization of incremental and radical NE does not appear to withstand scrutiny. Indeed, as I argue below, it is difficult to provide a plausible interpretation of such characterization. Concerning *incremental* NE, it remains unclear what exactly Camerer et al. mean when they prefigure the replacement of "as if assumptions that have never been well supported empirically" (2005, 10). After all, one might complain about the purported ad hocness or non-falsifiability of some 'as if' defences of rational choice theory. Yet, economists' 'as if' assumptions neither presuppose nor entail specific hypotheses concerning the neuro-psychological substrates of decisions. This, in turn, challenges NEs to explicate on what grounds empirical findings about these substrates would compel economists to replace their 'as if' assumptions. More generally, the question remains as to why economists should adopt the functional forms used by modellers whose methodological presuppositions and explanatory aims differ sharply from their own (see *Section 1*; see also *Section 3* for a discussion of NEs' calls to discriminate between economic models in terms of their relative fit with neural findings).

Regarding Camerer et al.'s counterfactual characterization of radical NE, one wonders whether it makes sense to ask how economics might have evolved if current neuroscience had influenced it from 'the start'. To be fair, one does not have to put forward a hyper-detailed counterfactual history of economic theory to defend Camerer et al.'s characterization of radical NE. Still, the limitations affecting our epistemic access to counterfactual developments in the history of economic theory significantly constrain the informativeness of such characterization. In particular, it is difficult to see on what basis we are to ascertain how economics might have evolved, had it been informed by insights and findings now available from neuroscience. Maybe we would have quasi-infallible, neurally informed economic models with tremendous predictive credentials. Or perhaps economists would have fallen prey of irredeemable confusion due to pan-explanatory *hubris*. In short, the range of possibilities is so wide that favouring one particular counterfactual scenario would appear to be quite arbitrary.

That said, let us try to articulate a more informative conceptualization of the incremental/radical divide. In his 2010 article, Fumagalli explicates the distinction between incremental and radical NE as follows. On the one hand, incremental NE proceeds on the assumption that economists' traditional constructs (e.g. preference relations, standard equilibrium concepts) provide a suitable basis for modelling people's decisions and relies on neuro-psychological findings to adjust or enrich specific economic *models*. On the other hand, radical NE challenges economists to modify or even replace their traditional constructs and aims to implement substantial changes in economic *theory*. Two remarks about this characterization of incremental and radical NE are in order.

First, the incremental/radical divide is best depicted not so much as an all-or-nothing dichotomy, but rather as a continuum along which intermediate positions can be differentiated. To be sure, whether one takes the accumulation of incremental modifications to constitute radical changes may depend on how she conceives of the relationship between economic models and economic theory.

Nonetheless, one may draw on entrenched accounts of the relationship between models and theories to argue that the divide between incremental and radical NE is more plausibly characterized as a matter of degree rather than an all-or-nothing affair (see e.g. Suppe, 1989, 3-37, on distinct variants of the semantic view of scientific theories, which regards such theories as collections of models).[5]

And second, Fumagalli's contrast between incremental and radical NE cuts across other informative categorizations of NEs' contributions. To see this, suppose you wanted to classify NEs' contributions in terms of how pervasively they influence economists' modelling practices. On the one hand, an incremental contribution may have considerable influence on such practices. For instance, NEs have recently succeeded in predicting individuals' decisions across a range of short-term, stimulus-bound rewards by monitoring the activation patterns of specific neural areas (see e.g. Levy and Glimcher, 2012). These studies do not currently yield predictions over sufficiently extended temporal horizons to be valuable to economists (Fumagalli, 2014). However, they pave the way for incremental NE contributions that could in principle prompt far-reaching modifications in economic modelling (e.g. think of NE models that reliably predicted decisions between long-term, non-stimulus-bound choice options). On the other hand, a radical contribution might have limited impact on economists' modelling practices. By way of illustration, consider recent attempts to 'bridge the gap' between neuroscience and economics by linking neuroscientific measurement protocols, observed behavioural responses, and formal choice theory through rigorous axiomatic statements.

In a pioneering contribution, Caplin and Dean (2008) develop an axiomatic model of reward valuation, which relates both experienced and predicted rewards to observable dopaminergic activations via a set of axiomatic statements. This model aims to provide an axiomatic foundation for the so-called 'reward prediction error' hypothesis, which states that dopaminergic activations encode the difference between experienced and predicted rewards (reward prediction error, RPE). Before this contribution, several studies had associated specific RPE models with fMRI measurements of neural activity in dopamine target areas (see e.g. Rutledge, 2010, for a review). Caplin and Dean (2008) move beyond these studies by identifying a set of necessary and sufficient conditions that neural activations from any brain area must satisfy to represent a RPE signal. Building on this work, Caplin et al. (2010) present the first application of an axiomatic representation theorem to test the reward prediction error hypothesis using neurobiological data. In their view, such application "rigorously tests and confirms [previous] evidence indicative of [such hypothesis] in fMRI data" (ibid., 951).

Caplin et al. provide neuroscientists interested in the neural underpinnings of belief formation and reward value learning with an innovative method to test entire classes of RPE functional forms. Moreover, their work illustrates how economists' axiomatic approach may be applied to model various phenomena besides those traditionally targeted by economists. Even so, Caplin et al. only demonstrate that their axioms of interest are satisfied at the aggregate level rather than for each experimental subject, and rely on some questionable assumptions (see e.g. Camerer,

[5] Fumagalli (2010) occasionally contrasts incremental and radical *NEs*, as opposed to incremental and radical *NE*. Here I focus on the latter contrast, since several authors provide both incremental and radical contributions in their studies and are therefore hard to categorize as either incremental or radical NEs.

2013, on the assumption that predicted rewards remain fixed based on stated probabilities, rather than being adjusted along a time path). Furthermore, the investigated dopaminergic activations relate to choice less directly than the workings of areas targeted by recent studies of the neural substrates of decisions (see e.g. Bartra et al., 2013). To put it differently, Caplin et al.'s results do not directly bear on economic modelling and do "not immediately advance our understanding of choice" (2010, 953).

At this stage, a proponent of NE may concede that NEs' contributions currently have limited *direct* relevance for the economic modelling of choice. At the same time, she might insist that such contributions prompt major *indirect* changes in economic modelling through their influence on other disciplines. To give one example, Padoa-Schioppa infers that neuroscience can contribute to economics from the alleged facts that "neuroscience can contribute to psychology, and that psychology can contribute to economics" (2008, 450-1). This claim points to a more sophisticated conceptualization of the relations that putatively hold between NE's parent disciplines than other NEs' assertions (see *Section 4*). However, it is an open question whether the vague notion of 'contribution' on which Padoa-Schioppa's claim is premised grounds an informative transitive inference. That is to say, while NEs have hitherto made limited contributions having direct relevance for economic modelling, their indirect contributions seem exceedingly speculative to license the enthusiasm surrounding NE research. To be fair, some NEs' exaggerations may be plausibly ascribed to the rhetoric of the discipline (see e.g. Mäki, 2010) and are expectable in light of NE's early stage of development (see e.g. Vromen, 2010). Yet, the point remains that even leading NEs significantly overstate the evidential and explanatory relevance of their findings for the economic modelling of choice.

(3) Do NEs' findings advance *model selection* in economics?

During a trip to Murano (Venice), you visit a glass craftsman in order to buy a present for your *fiancé*. As it happens, you are holding an expensive colourful vase of glass, and the vase falls out of your hand. A wide variety of models could be used to represent your attempt to catch the vase before it breaks on the ground. For example, your attempt may be modelled as if you were trying to minimize purely monetary losses, the acoustic noise resulting from the vase destruction, the sheer number of items that will predictably lie on the shop's floor, and so on. As this example suggests, one may employ many different models to represent observed choice behaviour. However, the mere fact that multiple models are compatible with observed choice behaviour does not imply that such models are empirically equivalent, i.e. that no behavioural, psychological, neural, etc. findings can discriminate between them.

What findings should economists employ to discriminate between their models of choice? Several NEs regard relative fit with the available neural findings as a prescriptive criterion for model selection in economics. The idea is to single out, among behaviourally equivalent models of choice, those that fit the available neural evidence best. Their reasoning goes as follows. Economists are often unable to discriminate between their models of choice on the sole basis of observed

decisions. Moreover, they frequently rely on "competing models that are either not tested, or if tested often explain the data equally well" (Vercoe and Zak, 2010, 133). Fortunately, neural findings enable economists to restrict the set of models that are compatible with the available evidence (see e.g. Rustichini, 2009). Hence, economists should discriminate between their models of choice in terms of relative fit with neural findings.[6]

Despite its apparent plausibility, this reasoning does not provide economists with convincing grounds to discriminate between their models of choice in terms of relative fit with neural findings. I shall explicate and discuss two reasons in support of this criticism in points *i* and *ii* below. The first reason is that neural findings rarely enable economists to restrict the set of choice models compatible with the available evidence to a significant extent. The second reason relates to NEs' failure to show that economists should discriminate between their models of choice in terms of relative fit with neural findings in cases where such findings and other disciplines' findings yield contrasting indications concerning which models fit the available evidence best.

i) NEs often claim that neural findings *significantly restrict* the set of choice models compatible with the available evidence. For example, after contending that economists provide various axiomatic systems consistent with observed decisions, Camerer alleges that "neural tests could winnow a crowded field of possible [models] down to the more plausible [candidates]" (2008, 47). Similarly, Rustichini complains that economists lack effective strategies for model selection and asserts that neural findings can play a "fundamental role" in "pruning the multiplicity of models [and making] them closer to the hard experimental test" (2009, 58).

Neural findings can in principle help choice modellers to restrict the set of models compatible with the available evidence. For instance, as noted by Glimcher (2011, ch.12), one might take the alleged fact that most sensory encoding of reward values is reference-dependent to favour prospect theory - which posits agents with a reference-dependent utility function (see e.g. Kahneman and Tversky, 1979) - over standard expected utility theory. Even so, neural findings rarely enable economists to overcome their putative problems of evidential underconstraint. To illustrate this, let us consider one class of economic models where neural findings are frequently claimed to facilitate model selection, namely multiple-self models.

Multiple-self models represent agents' decisions as the outcome of the interactions of various types of sub-personal entities (e.g. psychological processes, neural populations). A wide range of multiple-self models have been proposed in the economic literature. For example, some represent choices as the solution of a bargaining game among sub-personal agents with conflicting utility functions (see e.g. Benhabib and Bisin, 2005). Others, instead, characterize decisions as outcomes of the interplays between a sequence of short-run impulsive selves and a long-run

[6] Neural findings are claimed to serve discriminatory purposes in various disciplines besides economics (see e.g. Kable, 2011, on the opportunity to combine brain-imaging and brain-stimulation findings to discriminate between psychological theories). Here I focus on the use of such findings for model selection in economics. In doing so, I employ the expression 'available evidence' broadly to indicate the set of behavioural, psychological, neural, etc. findings that are taken to bear on the merits of the examined models unless I mention a specific subset of those findings.

patient self (see e.g. Fudenberg and Levine, 2006). NEs often conjecture that neural findings help economists to discriminate between multiple-self models by associating the sub-personal entities posited by these models with specific neural areas. Yet, it is doubtful that neural findings aptly serve this purpose. For *in primis*, economists make no explicit presuppositions as to whether the sub-personal entities posited by multiple-self models map onto particular neural counterparts (see e.g. Harrison, 2008). And second, severe evidential concerns affect NEs' intended applications of neural findings for model selection purposes in economics.

By way of illustration, consider leading NEs' debate as to whether multiple selves map onto anatomically separate neural populations in individuals' intertemporal choices. In a series of experiments, McClure et al. (2004 and 2007) examine the neural correlates of individuals' choices between pairs of monetary rewards available at distinct points in time. After observing that the ventral striatum, the medial orbitofrontal cortex, and the medial prefrontal cortex undergo differential activations during choice tasks involving temporally proximate versus more distant rewards, they infer that anatomically separate neural systems respectively value immediate and delayed monetary rewards. McClure et al.'s findings have been taken to provide convincing evidence in favour of a two-component model of intertemporal choice (see e.g. Carter et al., 2010). However, the reported evidence demonstrates neither that the examined areas discount rewards at different rates in each subject nor that these areas' discount rates differ significantly from the discount rates revealed by each subject's choices. Moreover, comparative NE studies document that a single-parameter hyperbolic discounting function fits observed neural activations better than McClure et al.'s two-component model (see e.g. Kable and Glimcher, 2007). Hence, it remains an open question whether the areas indicated by McClure et al. constitute plausible candidate neural referents for the multiple selves hypothesized to model agents' intertemporal choices.

Analogous considerations apply to other influential proposals to employ neural findings to discriminate between standard economic models. To see this, consider recent applications of NE studies of the neural substrates of reward valuation. Various NEs claim to have identified anatomically delimited neural areas that encode the subjective values of individuals' choice options on a common neural scale (see e.g. Levy and Glimcher, 2012). Some advocate measuring these areas' activations to construct "an independent measure of [subjective] value, in principle dissociable from choices", so as to render the claim that choices maximize subjective values "falsifiable" (Padoa-Schioppa, 2011, 335). In their view, building a neural measure of subjective value would constitute a major advancement in economic modelling, since subjective values cannot be measured independently of choices on the basis of choice data alone. Regrettably, neither constructing a neural measure of subjective value nor establishing under what circumstances choices maximize the value of such measure bears directly on the merits of economists' models. For these models make no assumption regarding what internal value function individuals actually use when assigning subjective values to particular choice options (see *Section 1*).

ii) Suppose facing a situation where the available neural evidence enables economists to discriminate between their models of choice, in the sense that it yields precise indications as to which of these models fit the collected neural findings best. Relative fit with neural findings is just one among *several* criteria for

discriminating between economic models of choice. Moreover, relative fit with neural findings and the modelling desiderata valued by economists (e.g. explanatory relevance) often make *conflicting* demands on modellers (see Fumagalli, 2011). Hence, some cogent reason is needed to license the claim that economists should discriminate between their models of choice in terms of relative fit with neural findings as opposed to some other criterion such as compatibility with other disciplines' findings (e.g. observed decisions, hedonic satisfaction reports, etc.).

A proponent of NE may answer that relative fit with neural findings and compatibility with other disciplines' findings are complementary - rather than competing - criteria for model selection in economics. The thought would be that neural findings provide economists with an additional source of evidence besides other disciplines' findings that is worth having on distinct metrics of informational value. Now, NEs frequently urge other economists to build models whose assumptions and implications are consistent with findings from multiple disciplines (see e.g. Glimcher and Rustichini, 2004). Furthermore, various authors advocate triangulating findings from different disciplines in order to improve specific models of choice (see e.g. Kable, 2011). Nonetheless, these contributions do not substantiate the claim that economists should employ neural findings to discriminate between their models of choice (see point *i* above). In particular, NEs' calls for this claim appear to face the following dilemma.

On the one hand, neural findings and other disciplines' findings may yield equivalent indications concerning which models fit the available evidence best. In this case, neural findings could increase one's confidence in other disciplines' findings, but do not provide economists with discriminatory insights beyond those yielded by such disciplines. This, coupled with the fact that modellers can rarely obtain accurate and reliable neural findings in situations where they lack access to behavioural and psychological findings, provides a reason to doubt that neural findings significantly advance model selection in economics. On the other hand, neural findings and other disciplines' findings may yield contrasting indications regarding which models fit the available evidence best. In this case, some cogent reason is needed to license the claim that economists should discriminate between their models of choice in terms of relative fit with neural - as opposed to other disciplines' - findings. Unfortunately, NEs have not offered yet precise and plausible criteria for dealing with cases where neural findings and other disciplines' findings yield contrasting indications concerning which models fit the available evidence best.

To be sure, some NEs manifest both descriptive and prescriptive ambitions, and maintain that economists should discriminate between their models of choice in terms of relative fit with neural findings. However, even leading NEs seem to overestimate the prescriptive implications of their findings for other economists. To see this, consider Glimcher et al.'s conjecture that by combining economic and neuroscientific findings, NE will develop "a methodology for reconciling prescriptive and descriptive economics" (2005, 214). NE findings may provide informative insights about the causal underpinnings of decisions, yet do not have direct implications regarding what people ought to choose in specific situations. Similarly, claiming that ideally choice models should be "tested simultaneously at the neural, psychological, and economic levels of analysis" (Glimcher, 2011, 132)

falls short of establishing that economists should discriminate between their models of choice in terms of relative fit with neural findings. For this regulative ideal does not specify how modellers should deal with cases where neural findings and other disciplines' findings conflict. Moreover, one may endorse such regulative ideal without being committed to regard compatibility with findings from multiple disciplines as a prescriptive criterion for model selection in economics. Indeed, requiring that economic models be compatible with all the available behavioural, psychological and neural findings would often be needlessly restrictive. For economists can frequently achieve their predictive and explanatory goals by building models whose assumptions neglect or even contradict specific subsets of these findings (see e.g. Fumagalli, 2014, on neural findings).

(4) What disciplines provide the *basic constructs* for NEs' accounts of choice?

The proponents of NE commonly emphasize that developing informative accounts of choice requires one to integrate findings and modelling tools across NE's parent disciplines (see e.g. McCabe, 2008). Nonetheless, NEs hold dissimilar views concerning what disciplines provide the *basic constructs* for their interdisciplinary accounts of choice. For instance, Dayan and Daw allege that decision theoretic concepts "permeate experiments and computational models in [both] psychology and neuroscience" and have "direct psychological and neural instantiations" (2008, 429 and 450). For their part, Camerer et al. assert that "the traditional economic account of behavior, which assumes that humans act so as to maximally satisfy their preferences, starts in the middle […] of the neuroscience account" (2005, 27). Still differently, Glimcher et al. maintain that "ultimately, economics is a biological science" (2005, 254), and Zak and Denzau contend that "findings in the biological sciences need to be incorporated directly into economics if the discipline is to continue to produce relevant insights into human behavior" (2001, 32).

The aforementioned assertions point to rather different conceptualizations of the intertheoretic relations that purportedly hold between NE's parent disciplines. *Prima facie*, this diversity might seem an unproblematic or even welcome indication of pluralism on NEs' part. After all - the thought would be - there is widespread disagreement as to how exactly NE's parent disciplines are related, and it would be unnecessarily demanding to require NEs to reach consensus on this issue. Furthermore, different disciplines may jointly contribute to providing the basic constructs for NEs' accounts of choice. In this Section, I inspect NEs' assertions and argue that these assertions are premised on contentious conceptualizations of the intertheoretic relations between distinct decision sciences. I then focus on prominent attempts to reduce economists' theoretical constructs to neural constructs and argue that such attempts rest on exceedingly speculative assumptions to effect the reductive accomplishments they envisage.

Consider first Camerer et al.'s claim that economists' account of behavior "starts in the middle […] of the neuroscience account" (2005, 27). This contention would be hardly informative if it restated the platitude that individuals' decisions are preceded by neural events that are often "inaccessible to consciousness" (ibid., 31). For neuroscientists have been repeating for decades that specific neural regions frequently activate before the moment agents report making particular decisions

(see e.g. Libet, 1983). Conversely, Camerer et al.'s claim would strain credulity, if it were meant to suggest that NEs have the means to identify the basic physical determinants of individuals' decisions. For neural events and processes result from the interactions of micro-physical constituents operating at more fine-grained spatio-temporal scales than those investigated by NEs.

Similar concerns arise regarding NEs' assertions about the purported biological character of economics. At first glance, it might seem that economics, defined as the study of human behaviour as a relationship between ends and scarce means having alternative uses (Robbins, 1932), has significant conceptual affinities with biology (see e.g. Marshall, 1890 [1961], book 4). This, however, falls short of licensing Glimcher et al.'s claim that "ultimately, economics is a biological science" (2005, 254). In particular, the hitherto proposed interpretations of this claim are vulnerable to the following criticism. On the one hand, interpreting the term 'biological' broadly - so as to suggest that economic models represent the behaviour of some generic living organisms - renders the claim trivial.[7] On the other hand, narrower interpretations of such term fail to clarify in what sense economics constitutes a biological discipline and what implications this has for the economic modelling of choice.[8] This does not exclude that NEs might find some plausible and precise conceptualization of 'biological' that avoids or circumvents the outlined criticism. Yet, it challenges those NEs who regard economics as a biological science to articulate such conceptualization.

Zak and Denzau's (2001, 32) contention that the advancement of economics is conditional upon the direct incorporation of biological findings is even more problematic. To be fair, various authors emphasize the relevance of biological findings for understanding economic phenomena (see e.g. Alchian, 1950, on how evolutionary thinking can shed light on the influences of competitive dynamics on agents' decisions). Furthermore, biological findings may in principle help economists to identify situations where their models' predictions are more likely to hold (see e.g. Satz and Ferejohn, 1994, on rational choice theory) and develop more explanatory models of choice (see e.g. Vromen, 2007, on the bioeconomic research program). This, however, by no means implies that the advancement of economics is conditional upon the direct incorporation of those findings. Indeed, it is hard to think of biological findings whose direct incorporation is necessary for the advancement of economic modelling and theorizing.

Faced with the above criticisms, a proponent of NE may concede that several authors put forward disputable *claims* concerning the intertheoretic relations

[7] Some authors (e.g. Ross et al., 2008, viii) claim that standard economic models do not rest on explicit assumptions as to what sort of entities (e.g. human individuals, firms, neural areas) representative agents supposedly map onto. Yet, even those authors grant that economic modellers are ultimately concerned with living organisms of some sort (see e.g. Ross, 2009; see also Davis, 2009, for a discussion).

[8] To give one example, Rosenberg infers that economics is a biological science on the sole ground that economists study "the causes and effects of the behavior of members of [the *Homo sapiens*] species" (2009, 60). The observation that economists investigate the causes and effects of humans' behaviour does not *per se* make economics a biological science in an interesting sense. Moreover, Rosenberg puts forward questionable claims in explicating the implications of the alleged biological character of economics. In particular, he nowhere substantiates his claim that "almost everything mysterious and problematical […] about economics is resolved once we understand economics as a biological science" (2009, 59).

between NE's parent disciplines. At the same time, she may rebut that NEs have made *substantial advances* in their attempts to link theoretical constructs across such disciplines. To assess this rebuttal, I shall focus on what many regard as the most advanced NE attempts to link economic and neural constructs, namely Glimcher et al.'s works on a partial reduction of decision utility to neural utility.

Over the last few years, NE studies have provided increasingly detailed evidence concerning what variables are encoded in the human brain while individuals make decisions. These studies (e.g. Kable and Glimcher, 2009) integrate brain-imaging data with single-neuron measurements in non-human primates to document systematic correlations between individuals' decisions and BOLD signals measured in areas such as the ventral striatum (VS) and the ventromedial prefrontal cortex (VMPFC). Building on these findings, various NEs contend that VS and VMPFC constitute the core of a neural valuation system for value-based learning and decision-making (see e.g. Levy and Glimcher, 2012). In particular, Glimcher conjectures that VS and VMPFC "contain all the neurons we require to extract [subjective values] for any object" (2009, 509).

If correct, Glimcher's conjecture could have momentous implications for choice modellers, since it may allow them to develop NE models that reliably predict choices even when expected utility theory fails to do so (Glimcher, 2011, ch.12-15). Moreover, it would provide a powerful rationale for modelling agents' utility functions as integral parts of a neural mechanism for choice, rather than purely 'as if' theoretical constructs. The idea is that while traditional economic theory only claims that decision makers choose 'as if' they used a common currency for valuing options, subjective value representations are in fact computed and integrated at the neural level during choice. This, in turn, would constitute a crucial step in the development of a partial reduction of decision utility to neural utility (see e.g. Vromen, 2012).

Unfortunately, there are both empirical and conceptual reasons to doubt Glimcher's conjecture that VS and VMPFC activations are constitutive of reward valuation at the whole-individual level. For example, the available evidence suggests that VS and VMPFC compute and integrate subjective values in situations where individuals face short-term stimulus-bound rewards. Yet, in spite of recent attempts to extend such findings to non-stimulus-bound valuations (see e.g. Cooper et al., 2013), we lack convincing evidence that the neural areas identified by Glimcher determine choice when it comes to the long-term non-stimulus-bound choices that interest economists (see e.g. Camerer, 2013, on choices that involve competition between habitual and goal-directed valuation systems). Furthermore, various studies document that other neural areas besides those indicated by Glimcher contribute to valuation in reward-guided behaviour (see e.g. Noonan et al., 2011, on complementary valuation processes implemented in different frontal cortical areas).

These findings do not exclude that NEs may identify statistically significant correlations between individuals' decisions and specific areas' activation patterns in increasingly sophisticated choice tasks. However, taken together, they cast serious doubt on hitherto proposed reductions of decision utility to neural utility. In particular, they strongly support the supposition that decision utility may routinely diverge from NEs' measures of subjective reward value. For several factors that

can be only partly accounted for in terms of neural computations of reward values (e.g. individuals' ethical commitments, contextual variations in institutional constraints) often mediate the influence of neural activations on choices in real-life situations. In these situations, too many and overly speculative inferential steps are required to reduce decision utility to NEs' measures of subjective reward value. As Ross puts it, "brains must always do something to produce behavior that implements choices; but this may not generally, let alone always, be direct neural computation of comparative reward values" (2011, 306).

(5) Does NE prompt a revolutionary expansion of the *evidential base* of economic theory?

Revolutionary scientific change has been the focus of heated methodological debates both in general philosophy of science (see e.g. Kuhn, 1962, and Lakatos, 1970) and in specific disciplines (see e.g. Baumberger, 1977, and Blaug, 1975, on economic theory). The proponents of NE frequently speak of implementing revolutionary modifications in NE's parent disciplines (see e.g. Shiller, 2011). In this last Section, I examine one respect in which NEs have claimed to foster such modifications. More specifically, I shall critically assess the thesis that NEs' contributions prompt a revolutionary expansion of the *evidential base* of economic theory, i.e. the set of explanatory variables that inform and are to figure into such theory. The idea is that NEs can observe variables that "are considered inherently unobservable" by other economists (Camerer, 2008, 45), and that NE's "largest payoff […] may come from pointing to biological variables which have a large influence on behaviour and are underweighted or ignored in [economic] theory" (Camerer, 2007, C35).[9]

To assess the cogency of these claims, let us distinguish the following three progressively broader conceptualizations of the evidential base of economic theory. On a first conceptualization, this base includes exclusively observable choices. This view has been advocated by a number of authors, ranging from early proponents of the revealed preference approach to some contemporary researchers (e.g. Gul and Pesendorfer, 2008). A more inclusive conceptualization holds that both observable choices and psychological variables (e.g. hedonic satisfaction reports) belong to the evidential base of economic theory. This view has been endorsed by several behavioural economists (e.g. Simon, 1955, and Kahneman and Tversky, 1979), who urged other economists to combine behavioural and psychological findings in constructing and assessing their models. According to a third, even broader conceptualization, the evidential base of economic theory comprises not just observable choices and psychological variables, but also neural variables. In this perspective, NEs' calls to incorporate neuro-psychological

[9] NE has been claimed to foster revolutionary changes in other respects besides the one examined in this Section. For instance, Glimcher and Rustichini (2004) emphasize the *scope* of NEs' proposed revolution, which supposedly encompasses the accounts of choice respectively developed by economists, psychologists, and neuroscientists. For his part, Glimcher (2011, ch.6) insists on the *depth* of NEs' purported revolution, which allegedly fosters the reduction of some basic constructs employed in NE's parent disciplines. I gloss over these considerations here, since I already commented on them in *Sections 1* and *4*.

variables into economic models of choice may be seen as the most recent wave of a long-lasting process of expansion of the evidential base of economic theory.[10]

Now, current brain-imaging and brain-stimulation tools do enable NEs to measure and causally manipulate several variables besides observed choices (see e.g. Camerer, 2008). This, however, falls short of indicating that NEs' contributions foster a significant - not to say revolutionary - expansion of the evidential base of economic theory. To be sure, I am not rehearsing Gul and Pesendorfer's dictum that "neuroscience evidence cannot refute economic models because the latter make no assumptions and draw no conclusions about the physiology of the brain" (2008, 4). For the mere fact that economists and neuroscientists respectively target different sets of explanatory variables does not *per se* preclude NEs from prompting a significant expansion of the evidential base of economic theory (e.g. the availability of increasingly detailed neural findings could in principle lead economists to adopt a broader conceptualization of such base). Still, NEs have not put forward convincing reasons to think that economists should include neural variables in the evidential base of economic theory. Let me explicate this concern.

Several leading NEs take their contributions to prompt a momentous expansion of the evidential base of economic theory. By way of illustration, Fehr and Camerer maintain that while economists "treat preferences and beliefs as impossible or difficult to observe directly, [NE] rejects the premise of unobservability" (2007, 419). Similarly, Levy and Glimcher contend that the available neural findings supply "a tool for measuring preferences neurobiologically" (2012, 1), and Camerer et al. allege that "the study of the brain and nervous system is beginning to allow direct measurement of thoughts and feelings" (2005, 10). *Prima facie*, these assertions might seem to be well supported by the ongoing advances in NEs' observational tools and experimental protocols. Regrettably, those assertions rest on disputable presuppositions concerning economists' constructs and the role such constructs play in economic theorizing. To illustrate this, let us focus on the notion of preference in contemporary rational choice theory.

Economic modellers and methodologists have debated at length about the merits of distinct interpretations of rational choice theory. Two such interpretations are frequently contrasted in the economic literature (see e.g. Guala, 2012, and Hands, 2012). On the one hand, we find a 'thin' behaviourist interpretation, according to which rational choice theory provides a purely formal representation of consistent behavioural patterns and makes no claim concerning the neuro-psychological underpinnings of choices. On the other hand, there is a 'thick' psychological interpretation, which takes rational choice theory to be grounded in a folk psychological conception of choices as the outcome of a process of instrumental reasoning. These behaviourist and psychological interpretations do not exhaust the

[10] This does not imply that NE is plausibly regarded as the mere continuation of behavioural and experimental economics with technologically more sophisticated instruments. I am not concerned here with discussing how NE differs from earlier research programs at the interface between economics and psychology. For present purposes, it suffices to note that NEs frequently criticize former works in behavioural and experimental economics. For instance, Glimcher et al. allege that most bounded rationality models "have little or no predictive power outside of their bounded domains" (2005, 214), and Glimcher contends that prospect theory "has too many interacting parameters to [be regarded as] a truly falsifiable theory" (2011, 120).

set of interpretations one may give to rational choice theory. Still, it is interesting to consider whether findings about the neural substrates of decisions supposedly expand the evidential base of such theory under each of those two interpretations.

NEs' findings do not directly expand the evidential base of rational choice theory when this theory is interpreted along 'thin' behaviourist lines. Indeed, the detractors of the 'thin' behaviourist interpretation commonly complain that adopting it severs rational choice theory from neuro-psychological research (see e.g. Guala, 2012). In this respect, adopting a 'thick' psychological interpretation might allow modellers to preserve continuity between rational choice theory and neuro-psychological research. Even so, one may endorse such interpretation, and yet deny that NEs' findings expand the evidential base of rational choice theory with regard to the notion of preference. After all, the availability of increasingly accurate measures of the neural correlates of preferences falls short of implying that such preferences are observable directly. More generally, neither the 'thin' behaviourist nor the 'thick' psychological interpretation of rational choice theory licenses the claim that preferences constitute a suitable target for direct neural observation (see e.g. Hausman, 2011).

At this stage, a proponent of NE might object that neural findings could foster an expansion of the evidential base of economic theory by prompting economists to refine their categorizations of economic phenomena or even employ different theoretical constructs. The idea would be to show that what was regarded as a unitary economic phenomenon in fact corresponds to distinct phenomena brought about by dissimilar neural mechanisms. Suppose that NEs succeed in demonstrating that particular economic phenomena (e.g. specific sets of intertemporal choices) are multiply realized at the neural level. This may be highly informative to choice modellers who investigate the neural substrates of decisions. However, it would neither mandate a refinement of economists' categorizations of economic phenomena nor foster a replacement of theoretical constructs such as time preferences and intertemporal discount rates. For economists define economic phenomena and constructs by reference to observable variations in individual and strategic behavioural patterns rather than to the workings of specific neural areas and processes (see e.g. Ross et al., 2008, 10). This does not exclude that NEs' findings might foster an expansion of the evidential base of economic theory. Still, it challenges NEs to put forward more convincing reasons to think that economists should include neural variables in the evidential base of economic theory.

Analogous remarks apply to the discovery that the neural substrates of putatively distinct economic phenomena overlap to a significant extent. By way of illustration, suppose that considerable overlaps are identified in the neural substrates of individuals' risk preferences, time preferences and social preferences. This finding may be evidentially and epistemically relevant for NE modellers. Yet, *pace* Fehr and Camerer (2007, 426), such finding does not appear to be especially "important" for other economists. In particular, it does not *per se* provide any convincing reason to think that economists should stop regarding these types of preferences as distinct. For the reasons why economists differentiate between those types of preferences relate not so much to the alleged lack of overlap in their neural substrates, but rather to observable variations in individual and strategic behavioural patterns. To put it differently, a lot of interesting work is under way at the interface between NE's parent disciplines. Still, NEs' findings are not shown to

foster a significant expansion of the evidential base of economic theory. This holds not just for some selected instances of self-promotional rhetoric on NEs' part, but also for NEs' claims concerning central economic constructs such as the notion of preference.

Conclusion

This paper argued for the following five theses on NE research: (1) NEs have made promising advances in integrating findings and modelling tools across NE's parent disciplines, but lack an adequate basis to provide a single unified model of choice spanning economics, psychology and neuroscience; (2) the proponents of NE justifiably resist some isolationist defences of standard economic theory, yet significantly overstate the evidential and explanatory relevance of NE findings for the economic modelling of choice; (3) NEs persuasively advocate the construction of economic models whose implications are compatible with other disciplines' findings, but do not presently provide convincing reasons to discriminate between economic models of choice in terms of relative fit with neural findings; (4) leading NEs have made innovative attempts to reduce economists' constructs to neural constructs, yet rely on exceedingly speculative assumptions to effect the reductive accomplishments they envisage; and (5) NEs expand the set of explanatory variables that may figure in interdisciplinary models of choice, but have hitherto failed to prompt a significant (not to say revolutionary) expansion of the evidential base of economic theory.

In commenting on the potential for success in early NE studies, several authors maintained that it is premature to judge NEs' achievements (Quartz, 2008, 466, and Smith, 2007, 313), that the ultimate test lies in NEs' future results (Bernheim, 2009, 38, and Schotter, 2008, 77), and the like. As argued in this paper, various empirical and conceptual issues still wait to be sorted out and clarified in contemporary NE research. This, however, licenses neither unreflective optimism about the prospects of NE nor wholesale methodological anarchy. On the contrary, it is precisely by addressing these issues that economic methodologists can provide an informative evaluation of NEs' contributions.

Acknowledgments

I thank J. McKenzie Alexander, Richard Bradley, John Davis, Wade Hands, Don Ross, and Jack Vromen for their comments on earlier versions of this paper. I also benefited from the observations of two anonymous referees and audiences at the University of Münster, the London School of Economics, and the Erasmus University of Rotterdam.

REFERENCES

Alchian, A. 1950. Uncertainty, Evolution, and Economic Theory. *Journal of Political Economy*, 58, 211-21.

Bartra, O., McGuire, J.T. and Kable, J.W. 2013. The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *NeuroImage*, 76, 412–427.

Baumberger, J. 1977. No Kuhnian Revolutions in Economics. *Journal of Economic Issues*, 11, 1-20.

Benhabib, J. and Bisin, A. 2005. Modeling internal commitment mechanisms and self-control: a neuroeconomics approach to consumption-saving decisions. *Games and Economic Behavior*, 52 (2), 460-92.

Bernheim, B.D. 2009. On the Potential of Neuroeconomics: A Critical (but Hopeful) Appraisal. *American Economic Journal: Microeconomics*, 1 (2), 1-41.

Blaug, M. 1975. Kuhn versus Lakatos, or paradigms versus research programmes in the history of economics. *History of Political Economy*, 7 (41), 399-433.

Camerer, C.F. 2008. The Case for Mindful Economics. In *The Foundations of Positive and Normative Economics. A Handbook*. Caplin, A. and Schotter, A. Eds. 43-69.

Camerer, C.F. 2013. A Review Essay about *Foundations of Neuroeconomic Analysis* by Glimcher, P. *Journal of Economic Literature*, 51(4), 1155-1182.

Camerer, C.F., Loewenstein, G. and Prelec, D. 2005. Neuroeconomics: how neuroscience can inform Economics. *Journal of Economic Literature*, 43 (1), 9-64.

Caplin, A. and Dean, M. 2008. Dopamine, reward prediction error, and economics. *Quarterly Journal of Economics*, 123 (2), 663–701.

Caplin, A., Dean, M., Glimcher, P.W. and Rutledge, R.B. 2010. Measuring beliefs and rewards: A neuroeconomic approach. *Quarterly Journal of Economics*, 125 (3), 923-960.

Carter, R. McKell, Justin R. Meyer, and Scott A. Huettel. 2010. Functional Neuroimaging of Intertemporal Choice Models: A Review. *Journal of Neuroscience, Psychology, and Economics*, 3(1), 27–45.

Cooper, N., Kable, J. W., Kim, B. K., and Zauberman, G. 2013. Brain activity in valuation regions while thinking about the future predicts individual discount rates. *The Journal of Neuroscience,* 33, 13150-13156.

Craver, C.F. and Alexandrova, A. 2008. No Revolution Necessary: Neural Mechanisms for Economics. *Economics and Philosophy*, 24, 381-406.

Davis, J.B. 2009. Competing Conceptions of the Individual in Recent Economics. In *The Oxford Handbook of Philosophy of Economics*, Ed. Kincaid, H. and Ross, D. Oxford University Press, 223-244.

Dayan, P. and Daw, N. 2008. Decision theory, reinforcement learning, and the brain. *Cognitive, Affective, & Behavioral Neuroscience*, 8 (4), 429-453.

Dorris, M.C. and Glimcher, P.W. 2004. Activity in Posterior Parietal Cortex is Correlated with the Relative Subjective Desirability of Action. *Neuron*, 44, 365–378.

Fehr, E. and Camerer, C.F. 2007. Social neuroeconomics: the neural circuitry of social preferences. *Trends in Cognitive Sciences*, 11 (10), 419-426.

Fehr, E. and Rangel, A. 2011. Neuroeconomic Foundations of Economic Choice - Recent Advances. *Journal of Economic Perspectives*, 25 (4), 3–30.

Fudenberg, D. and Levine, D. 2006. A Dual Self Model of Impulse Control. *American Economic Review*, 96, 1449-1476.

Fumagalli, R. 2010. The disunity of neuroeconomics: a methodological appraisal. *Journal of Economic Methodology*, 17 (2), 119-131.

Fumagalli, R. 2011. On the neural enrichment of economic models: tractability, trade-offs and multiple levels of description. *Biology and Philosophy*, 26 (5), 617-635.

Fumagalli, R. 2013. The Futile Search for True Utility. *Economics and Philosophy*, 29 (3), 325-347.

Fumagalli, R. 2014. Neural Findings and Economic Models: Why Brains have Limited Relevance for Economics. *Philosophy of the Social Sciences*, 44 (5), 606-629.

Glimcher, P.W. 2009. Choice: towards a standard back-pocket model. In P.W. Glimcher, C. Camerer, E. Fehr, and R. Poldrack (eds.). *Neuroeconomics: Decision Making and the Brain*. London: Elsevier, 503-521.

Glimcher, P.W. 2011. *Foundations of Neuroeconomic Analysis*. Oxford University Press.

Glimcher, P.W., Dorris, M.C. and Bayer, H.M. 2005. Physiological utility theory and the neuroeconomics of choice. *Games and Economic Behavior*, 52, 213-256.

Glimcher, P.W. and Rustichini, A. 2004. Neuroeconomics: The Consilience of Brain and Decision. *Science*, 306 (5695), 447-452.

Guala, F. 2012. Are Preferences for Real? Choice Theory, Folk Psychology, and the Hard Case for Commonsensible Realism. In *Economics for Real: Uskali Mäki and the Place of Truth in Economics*, A. Lehtinen et al. (Ed.). Routledge, 137-155.

Gul, F. and Pesendorfer, W. 2008. The case for mindless economics. In Caplin, A. and Schotter, A. Eds. *The foundations of positive and normative economics*. Oxford University Press, 1-40.

Hands, W.D. 2012. Realism, Commonsensibles, and Economics: The Case of Contemporary Revealed Preference Theory. In *Economics for Real: Uskali Mäki and the Place of Truth in Economics*, A. Lehtinen et al. (Ed.). Routledge, 156-178.

Harrison, G. 2008. Neuroeconomics: a Critical Reconsideration. *Economics and Philosophy*, 24, 303-344.

Harrison, G. and Ross, D. 2010. The methodologies of neuroeconomics. *Journal of Economic Methodology*, 17 (2), 185-196.

Hausman, D.M. 1992. *The Inexact and Separate Science of Economics*. Cambridge University Press.

Hausman, D.M. 2011. Mistakes about Preferences in the Social Sciences. *Philosophy of the Social Sciences*, 41, 3-25.

Kable, J.W. 2011. The cognitive neuroscience toolkit for the neuroeconomist: A functional overview. *Journal of Neuroscience, Psychology, and Economics*, 4 (2), 63–84.

Kable, J.W. and Glimcher, P.W. 2007. The neural correlates of subjective value during intertemporal choice. *Nature Neuroscience*, 10, 1625-1633.

Kable, J.W. and Glimcher, P.W. 2009. The Neurobiology of Decision: Consensus and Controversy. *Neuron*, 63 (6), 733-745.

Kahneman, D. 2003. A Psychological Perspective on Economics. *American Economic Review*, 93 (2), 162-168.

Kahneman, D. and Tversky, A. 1979. Prospect Theory. An analysis of decision under risk. *Econometrica*, 47 (2), 263-291.

Kuhn, T.S. 1962. *The Structure of Scientific Revolutions*. University of Chicago Press.

Lakatos, I. 1970. Falsification and the methodology of scientific research programmes. In I. Lakatos and A. Musgrave, eds. *Criticism and the growth of knowledge*. Cambridge University Press, 91-195.

Levy, D.J. and Glimcher, P.W. 2012. The root of all value: A neural common currency for choice. *Current Opinion in Neurobiology*, 22, 1-12.

Libet, B. 1983. Time of conscious intention to act in relation to onset of cerebral activity. *Brain*, 106, 23-42.

Loomes, G. and Sugden, R. 1982. Regret Theory: an Alternative Theory of Rational Choice under Uncertainty, *Economic Journal*, 92, 805-24.

Mäki, U. 1996. Two portraits of economics. *Journal of Economic Methodology*, 3 (1), 1-38.

Mäki, U. 2010. When economics meets neuroscience: hype and hope. *Journal of Economic Methodology*, 17 (2), 107-117.

Marshall, A. 1890 [1961]. *Principles of Economics*, 9th ed. London: Macmillan.

McCabe, K. 2008. Neuroeconomics and the Economic Sciences. *Economics and Philosophy*, 24, 345-368.

McClure, S.M., Ericson, K.M., Laibson, D.I., Loewenstein, G. and Cohen, J.D. 2007. Time Discounting for Primary Rewards. *Journal of Neuroscience*, 27 (21), 5796-5804.

McClure, S.M., Laibson, D.I., Loewenstein, G. and Cohen, J.D. 2004. Separate Neural Systems Value Immediate and Delayed Monetary Rewards. *Science*, 306, 503–507.

Mundale, J. and Bechtel, W. 1996. Integrating Neuroscience, Psychology, and Evolutionary Biology Through a Teleological Conception of Function. *Minds and Machines*, 6, 481-505.

Noonan, M.P., Mars, R.B. and Rushworth, M.F. 2011. Distinct roles of three frontal cortical areas in reward-guided behavior. *Journal of Neuroscience*, 31 (40), 14399-412.

Padoa-Schioppa, C. 2008. The syllogism of neuro-economics. *Economics and Philosophy*, 24, 449–457.

Padoa-Schioppa, C. 2011. Neurobiology of Economic Choice: A Good-Based Model. *Annual Review of Neuroscience*, 34, 333-359.

Piccinini, G. and Craver, C. 2011. Integrating psychology and neuroscience: functional analyses as mechanism sketches. *Synthese*, 183, 283-311.

Quartz, S.R. 2008. From cognitive science to cognitive neuroscience to neuroeconomics. *Economics and Philosophy*, 24, 459–471.

Rabin, M. 1998. Psychology and Economics. *Journal of Economic Literature*, 36 (1), 11-46.

Robbins, L. 1932. *An Essay on the Nature and Significance of Economic Science*. London: Macmillan.

Rosenberg, A. 2009. If Economics is a Science, What Kind of a Science is it? In *The Oxford Handbook of Philosophy of Economics*, Ed. Kincaid, H. and Ross, D. Oxford University Press, 55-67.

Ross, D. 2008. Two Styles of Neuroeconomics. *Economics and Philosophy*, 24, 473-483.

Ross, D. 2009. Integrating the Dynamics of Multiscale Economic Agency. In *The Oxford Handbook of Philosophy of Economics*, Ed. Kincaid, H. and Ross, D. Oxford University Press, 245-279.

Ross, D. 2011. Cognitive Variables and Parameters in Economic Models. In *Grounding Social Sciences in Cognitive Science*, Ed. R Sun. MIT Press, 287-314.

Ross, D., Sharp, C., Vuchinich, R. and Spurrett, D. 2008. *Midbrain Mutiny: The Picoeconomics and Neuroeconomics of Disordered Gambling*. MIT Press.

Rustichini, A. 2009. Is There a Method of Neuroeconomics? *American Economic Journal: Microeconomics*, 1 (2), 48-59.

Rutledge, R.B., Dean, M., Caplin, A. and Glimcher, P.W. 2010. Testing the Reward Prediction Error Hypothesis with an Axiomatic Model. *Journal of Neuroscience*, 30(40), 13525–36.

Satz, D. and Ferejohn, J. 1994. Rational Choice and Social Theory. *The Journal of Philosophy*, 91 (2), 71-87.

Schotter, A. 2008. What's so informative about choice? In *The Foundations of Positive and Normative Economics: A Handbook*, Ed. Caplin, A. and Schotter, A. Oxford University Press, Ch.3, 70-94.

Shiller, R.J. 2011. The Neuroeconomics Revolution. Available at: http://www.project-syndicate.org /commentary/the-neuroeconomics-revolution.

Simon, H.A. 1955. A Behavioral Model of Rational Choice. *Quarterly Journal of Economics*, 69 (1), 99-118.

Smith, V.L. 2007. *Rationality in Economics: Constructivist and Ecological Forms.* Cambridge University Press.

Suppe, F. 1989. *The Semantic Conception of Theories and Scientific Realism.* University of Chicago Press.

Vercoe, M. and Zak, P.J. 2010. Inductive modeling using causal studies in neuroeconomics: brains on drugs. *Journal of Economic Methodology*, 17 (2), 133-146.

Vromen, J. 2007. Neuroeconomics as a Natural Extension of Bioeconomics: The Shifting Scope of Standard Economic Theory. *Journal of Bioeconomics*, 9, 145–167.

Vromen, J. 2010. Where economics and neuroscience might meet. *Journal of Economic Methodology*, 17 (2), 171-183.

Vromen, J. 2011. Neuroeconomics: two camps gradually converging: what can economics gain from it? *International Review of Economics*, 58, 267–285.

Vromen, J. 2012. Review of *Foundations of Neuroeconomic Analysis*, P.W. Glimcher. Oxford University Press, 2010. *Economics and Philosophy*, 28 (1), 108-113.

Zak, P.J. and Denzau, A. 2001. Economics is an evolutionary science. In *Evolutionary approaches in the behavioral sciences: toward a better understanding of human nature*. Somit and Peterson, 31-65.